6/PRTS

10/018009
PCT/EP00/06227
JC07 Rec'd PCT/PTO 14 DEC 2001

GENES ENCODING ENZYMES IN THE BIOSYNTHESIS OF PIMARICIN AND THE APPLICATION THEREOF

Field of the invention

The invention relates to novel genes encoding enzymes which are fundamental in the biosynthesis of pimaricin. The invention further relates the application of said gene for modifying the biosynthesis of pimaricin. It also relates to the biosynthesis of new compounds.

Background of the invention

Polyketides, such as pimaricin (in the literature also referred to as natamycin, see for its structure Fig. 3A), form a large and highly diverse group of natural products. Members of the said group include compounds having antibacterial, antifungal, anticancer, antiparasitic and immunosuppressant activities.

Despite their structural diversity, these metabolites are believed to be synthesized by micro-organisms by a common pathway in which units derived from acetate, propionate or butyrate are condensed onto a growing chain by a polyketide synthase (PKS). The process resembles fatty acid biosynthesis, except that the β-keto function introduced at each elongation step may undergo all, part or none of a reductive cycle comprising β-ketoreduction, dehydration and enoylreduction. Structural variety of polyketides arises from the choice of monomers, the extent of β-ketoreduction and dehydration, and the stereochemistry of each chiral center. Yet further diversity is produced by functionalization of the polyketide chain by the action of glycosylases, methyltransferases and oxidative enzymes.

Modification of complex biomolecules, such as polyketides, is increasingly an important way of obtaining biologically active compounds with improved or altered properties. Currently, these modifications are usually introduced by chemical methods in a directed or random (e.g.

in combinatorial chemistry) manner. A drawback of these chemical methods is that they are often performed under relatively harsh conditions and furthermore, they lack selectivity and/or sensitivity. Particularly, in the case of complex biomolecules having multiple functionalized, reactive groups, precautions have to be taken in order to avoid undesired side reactions. These precautions include for instance the introduction of protective groups before a desired chemical conversion. Consequently two additional process steps are involved, as the protective groups must be removed afterwards.

Bioconversion of simple organic compounds, i.e. compounds with no or single reactive centers, has been known for some time and has been widely applied. Examples are the oxidation of long chain alkanes using alkane hydroxylation systems of *Pseudomonas*, and epoxidation of alkenes using enzyme systems from various micro-organisms. However, for the specific modifications required in the biosynthesis of complex molecules, for example, β-lactam antibiotics, polyketide antibiotics, anticancer agents, or peptide antibiotics, the large amounts of reactive groups present in those molecules are problematic for even the simplest treatments, such as hydrolysis of specific bonds. More complicated treatments frequently completely destroy the molecule.

Summary of the invention

The present invention is based on the identification and isolation of three genes which encode enzymes which facilitate specific oxidative conversions in the biosynthesis of pimaricin. The present invention thus provides the means to perform specific conversions in complex biomolecules, in particular in polyketides, without applying the harsh conditions often related to chemical modifications. The said conversions can be carried as part

of a biosynthesis of said biomolecules, for instance in micro-organisms.

Surprisingly, it has been found that the expression of polynucleotides of the invention in different micro-organisms, can lead to the biosynthesis of different biomolecules. It has further been found that expression of the said polynucleotides may be switched off (or knocked out) in *Streptomyces* which is usually used for the biosynthesis of pimaricin. In this embodiment, no pimaricin is produced by said *Streptomyces*, but instead a modified biomolecule is produced. In addition, it has been found that the polynucleotides may be overexpressed in *Streptomyces*, leading to an increase in the biosynthesis of pimaricin in the said *Streptomyces*.

According to the invention there is thus provided a polynucleotide comprising:

i) a nucleic acid sequence set out in SEQ ID NO: 5, 7 or 9 or a sequence complementary thereto; or

ii) a homologue or fragment of a sequence defined in i).

The invention also provides:

- a polypeptide encoded by a polynucleotide of the invention which is preferably isolated and/or purified;
- a polypeptide obtainable by a polynucleotide of the invention in a cell which is a *Streptomyces* (including e.g. *S.natalensis*) cell or a cell of a heterologous species
- a polypeptide comprising the amino acid sequence set out in SEQ ID NO: 6, 8 or 9 or a homologue or fragment thereof;
- a recombinant cell comprising at least one additional copy of a polynucleotide of the invention, wherein the cell naturally possesses at least one said polynucleotide;
- a recombinant cell, wherein a polynucleotide of the invention which naturally occurs in the cell has been inactivated;

- a recombinant cell comprising a polynucleotide according to the invention which polynucleotide does not naturally occur in that cell or where the polynucleotide is heterologous to that cell;
- a method for overexpressing a polynucleotide encoding a polypeptide according to the invention in *Streptomyces* cell which method comprises:
  - i) attaching a promoter sequence to the said polynucleotide;
  - ii) transferring the resulting promoter-polynucleotide complex into the said cell; and
  - iii) maintaining the resulting cell under conditions suitable for expression of the said polynucleotide;
- a method for inactivating a polynucleotide encoding a polypeptide according to the invention in a *Streptomyces* cell which method comprises disrupting the coding sequence of the said polynucleotide;
- a method for expressing a polynucleotide encoding a polypeptide according to the invention in a heterologous cell which method comprises:
  - i) attaching a promoter sequence to the said polynucleotide;
  - ii) transferring the resulting promoter-polynucleotide complex into the said cell; and
  - iii) maintaining the resulting cell under conditions suitable for expression of the said polynucleotide;
- a method for producing pimaricin which method comprises maintaining a recombinant cell according to the invention under conditions suitable for obtaining expression of the additional copy of a polynucleotide according to the invention and isolating the said pimaricin;
- a method for producing a biomolecule which method comprises maintaining a recombinant cell according to the invention under conditions which would be suitable for obtaining expression of the inactivated polynucleotide

had it not been inactivated and isolating the said biomolecule;

- a method for producing a biomolecule which method comprises maintaining a recombinant cell according to the invention under conditions suitable for obtaining expression of the polynucleotide which does not naturally occur in the cell and isolating the said biomolecule;
- a biomolecule obtainable by a method of the invention for producing a biomolecule;
- use of a recombinant cell of the invention in the production of pimaricin;
- use of a recombinant cell of the invention in the production of a biomolecule;
- a vector containing a polynucleotide of the invention which is capable of expressing a polypeptide of the invention;
- a cell harbouring a vector of the invention; and
- a method for producing a polypeptide of the invention, which method comprises maintaining a recombinant cell according to the invention under conditions suitable for obtaining expression of the polypeptide and isolating the said polypeptide.
- use of a isolated and/or purified polypeptide according to the invention for the oxidative modification of a methyl group of a suitable compound.

Brief description of the drawings

Figure 1: Physical map of part of the Pimaricin biosynthetic cluster.

Genes: locations of the genes encoding polyketide synthases and oxidative genes involved in Pimaricin production (not drawn to scale);

Probes: 0.7 indicates the location of the 0.7 kb fragment used to identify the extent of polyketide synthase encoding regions; 3.3 indicates the location of the 3.3 kb fragment used in polyketide synthase gene disruption;

Cosmids: sizes and numbers of available cosmids covering the chromosomal region encompassing the oxidative genes.

Figure 2: Detailed physical map of the chromosomal regions including the oxidative genes.

Figure 3A: Molecular structure of Pimaricin.

Figure 3B: Molecular structures of Pimaricin derivatives with a reduced oxidation state of C4 and C5 and/or the carboxyl group at C12.

Figure 4: Molecular structures of Amphotericin B and Nystatin

Figure 5: 5 illustrates the conversion of the triketide lactone to it oxidized form by the action of pORF1 and pORF2

Description of the sequence listings

SEQ ID 1 shows the nucleotide sequence and derived amino acid sequence of a first Pimaricin biosynthesis associated polyketide synthase gene

SEQ ID 2 shows the amino acid sequence of a first Pimaricin biosynthesis associated polyketide synthase

SEQ ID 3 shows the nucleotide sequence and derived amino acid sequence of a second Pimaricin biosynthesis associated polyketide synthase gene

SEQ ID 4 shows the amino acid sequence of a second Pimaricin biosynthesis associated polyketide synthase

SEQ ID 5 shows the nucleotide sequence and derived amino acid sequence of ORF1, an oxidative gene involved in Pimaricin biosynthesis

SEQ ID 6 shows the amino acid sequence of an oxidation enzyme pORF1 involved in Pimaricin biosynthesis

SEQ ID 7 shows the nucleotide sequence and derived amino

acid sequence of ORF2, an oxidative gene involved in Pimaricin biosynthesis

SEQ ID 8 shows the amino acid sequence of an oxidation enzyme pORF2 involved in Pimaricin biosynthesis

SEQ ID 9 shows the nucleotide sequence and derived amino acid sequence of ORF3, an oxidative gene involved in Pimaricin biosynthesis

SEQ ID 10 shows the amino acid sequence of an oxidation enzyme pORF3 involved in Pimaricin biosynthesis

SEQ ID 11 shows a synthetic oligonucleotide (forward primer) for isolation by PCR of the ermE promoter of *Saccharopolyspora erythraea*

SEQ ID 12 shows a synthetic oligonucleotide (reverse primer) for isolation by PCR of the ermE promoter of *Saccharopolyspora erythraea*

SEQ ID 13 shows a synthetic oligonucleotide (forward primer) for isolation by PCR of the N-terminal region of ORF1

SEQ ID 14 shows a synthetic oligonucleotide (reverse primer) for isolation by PCR of the N-terminal region of ORF1

Detailed description of the invention

Three open-reading frames (ORFs) were identified from the Pimaricin producing microorganism *Streptomyces natalensis*. The three ORFs are associated with polyketide synthese genes and each ORF has been shown to be essential for pimaricin biosynthesis.

The functionality of the Pimaricin PKS associated genes was initially pursued by comparing their derived amino acid sequences with those present in public databases like EMBL, Genbank, NBRF/PIR, or Swissprot.

Surprisingly, ORF1 appeared to resemble cholesterol oxidases from several Streptomyces species. The close association of ORF1 with the Pimaricin PKS suggests an oxidative step in Pimaricin tailoring. A methyloxidase encoding gene has not been observed previously in a polyketide biosynthesis gene cluster.

Based on similar analyses, ORF2 and ORF3 resemble cytochrome P450 dependent monooxygenases from various sources. With respect to the biosynthesis of bioactive compounds, P450 dependent monooxygenases have been identified before in association with polyketide gene clusters, e.g. in the Erythromycin and Rapamycin biosynthesis gene clusters. However, only in the Erythromycin case has a specific enzymatic action on Erythromycin precursor compounds been proven. Essentially all known cases of tailoring oxidation steps act on secondary carbon atoms (methylene groups). Oxidation of primary carbon atoms (methyl groups)to carboxylic acid function in polyketide biosynthesis, as has presently been found, is unprecedented. Nothing is known about the molecular basis of epoxide formation in polyketide products, though epoxides are present in a few known structures.

Thus, the invention provides a polynucleotide which comprises:

i) a nucleic acid sequence set out in SEQ ID NO: 5, 7, or 9 or a sequence complementary thereto; or

ii) a homologue or fragment of a sequence defined in i).

Polynucleotides of the invention may comprise DNA or RNA. The invention also provides double stranded polynucleotides comprising a polynucleotide of the invention and its complement.

Homologues of a nucleic acid sequence set out in SEQ ID NO: 5, 7 or 9 are polynucelotideds which do not share 100% sequence identity with a sequence set out in SEQ ID NO: 5, 7, or 9, but which do encode polypeptides having a similar enzyme activity to a polypeptide encoded by a nucleic acid sequence set out in SEQ ID NO: 5, 7 or 9. Thus a homolog of a polypeptide encoded by SEQ ID NO: 5 will typically encode a polypeptide which has methyl oxidase or methyloxidase-like activity. A homologue of a polypeptide encoded by SEQ ID NO: 7 or 9 will typically encode a

polynucleotide which has cytochrome P-450 monooxygenase activity or cytochrome P-450 monooxygenase-like activity. A homologue of the invention will generally have at least 90%, at least 95%, at least 98% or at least 99% sequence identity to the sequence of SEQ ID NO: 5, 7 or 9 over a region of at least 60, more preferably at least 100 contiguous nucleotides or most preferably over the full length of SEQ ID NO: 5, 7 or 9 (for determination of sequence identity see D.J. Lipman, W.R. Pearson. 1985. Science 227, p1435).

Any combination of the above mentioned degrees of sequence identity and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher sequence identity over longer lengths) being preferred. Thus, for example a polynucleotide which has at least 90% sequence identity over 60, forms one aspect of the invention, as does a polynucleotide which has at least 95% sequence identity over 100 nucleotides.

The sequence of SEQ ID NO: 5, 7 or 9 may be modified by nucleotide substitutions, for example from 1, 2 or 3 to 10 or 25 substitutions. The polynucleotide of SEQ ID NO: 5, 7 or 9 may alternatively or additionally be modified by one or more insertions and/or deletions and/or by an extension at either or both ends. The modified polynucleotide generally encodes a polypeptide which has methyl oxidase or cytochrome P-450 monooxygenase activity.
Degenerate substitutions may be made and/or substitutions may be made which would result in a conservative amino acid substitution when the modified sequence is translated, for example as shown in the Table below.

Polynucleotides of the invention include fragments of a sequence set out in SEQ ID NO: 5, 7 or 9. Thus, polynucleotides of the invention may be used as a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labeled with a revealing label by conventional means using radioactive or non-radioactive

labels, or the polynucleotides may be cloned into vectors (M.A. Innis et al..1990. PCR Protocols, Academic Press Inc).

Such primers, probes and other fragments will preferably be at least 10, preferably at least 15 or at least 20, for example at least 25, at least 30 or at least 40 nucleotides in length. They will typically be up to 40, 50, 60, 70, 100, or 150 nucleotides in length. Probes and fragments can be longer than 150 nucleotides in length, for example up to 200, 300, 400, 500, 600, 700 nucleotides in length, or even up to a few nucleotides, such as five or ten nucleotides, short of the full length of the sequence of SEQ ID NO: 5, 7 or 9.

Polynucleotides such as DNA polynucleotide and primers according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques. The polynucleotides are typically provided in isolated and/or purified form.

In general, primers will be produced by synthetic means, involving a stepwise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using PCR (polymerase chain reaction) cloning techniques.

Although in general the techniques mentioned herein are well known in the art, reference may be made in particular to Sambrook *et al*, 1989, Molecular Cloning: a laboratory manual.

A polypeptide of the invention comprises the amino acid sequence set out in SEQ ID NO: 6, 8 or 10 or a substantially homologous sequence, or a fragment of the said sequences and typically has methyl oxidase or cytochrome P-450 monooxygenase activity. In general, the naturally

occurring amino acid sequence shown in SEQ ID NO: 6, 8 or 10 is preferred.

A polypeptide of the invention may comprise:

(a) the polypeptide sequence of SEQ ID NO: 2, 4, 6, 8, 10 or 12; or

(b) a homologue or fragment thereof.

A homologue may occur naturally, for example, in a bacterium and will function in a substantially similar manner to the protein of SEQ ID NO: 6, 8 or 10, for example it acts as a methyl oxidase in the case of a homologue of SEQ ID NO: 6 or a cytochrome P-450 monooxygenase in the case of a homologue of SEQ ID NO: 8 or 10.

Homologues can be obtained by following the procedures described herein for the production of the polypeptides of SEQ ID NO: 6, 8 or 10 and performing such procedures on a suitable cell source e.g. a bacterial cell. It will also be possible to use a probe as defined above to probe libraries made from bacterial cells in order to obtain clones encoding homologues. The clones can be manipulated by conventional techniques to generate a polypeptide of the invention which can then be produced by recombinant or synthetic techniques known *per se*.

A homologue of a polypeptide of the invention preferably has at least 80% sequence identity to the protein of SEQ ID NO: 6, 8 or 10, or more preferably at least 90%, at least 95%, at least 97% or at least 99% sequence identity thereto over a region of at least at least 40, preferably at least 60, for instance at least 100 contiguous amino acids or over the full length of SEQ ID NO: 6, 8 or 10.

The sequence of the polypeptide of SEQ ID NO: 6, 8 or 10 and of homologues can thus be modified to provide polypeptides of the invention. Amino acid substitutions may be made, for example from 1, 2 or 3 to 10 or 20. substitutions. The modified polypeptide generally retains activity as a methyl oxidase or cytochrome P-450 monooxygenase. Conservative substitutions may be made, for

example according to the following Table. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other.

ALIPHATIC	Non-polar	G A P
		I L V
	Polar-uncharged	C S T M
		N Q
	Polar-charged	D E
		K R
AROMATIC		H F W Y

Polypeptides of the invention also include fragments of the above-mentioned full length polypeptides. Such fragments typically retain activity as a methyl oxidase or cytochrome P-450 monooxygenase.

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus, in a further embodiment, the invention provides a method of making polypeptides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell.

Preferably, a polynucleotide of the invention in a vector is operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence, such as a promoter, "operably linked" to a coding sequence is positioned in such a way that

expression of the coding sequence is achieved under conditions compatible with the regulatory sequence.

Vectors of the invention may be transformed into a suitable host cell to provide for expression of a polypeptide of the invention. Thus, the invention provides a process for preparing a polypeptide according to the invention which comprises cultivating a host cell transformed or transfected with an expression vector encoding the polypeptide and recovering the polypeptide.

Each of the genes ORF1, ORF2 and ORF3 can be used for various purposes separately or in combination. This will be discussed in detail below.

Targeted inactivation of one or more of the present genes, e.g. through marker insertion or replacement with a non-functional gene equivalent, interferes with at least one (oxidation) step in the Pimaricin biosynthetic route. This results in the production of modified Pimaricin molecules characterized by a different oxidative state. For example, molecules can be created lacking the epoxide function at carbons C4 and C5 , or molecules with a modified oxidation state of the carboxyl group at C12 resulting in an aldehyde, alcohol, or methyl group at this position.

Disruption of chromosomally encoded genes can be accomplished by gene replacement strategies. Gene replacement is preferably carried out using suicide plasmid vectors or defective phage vectors carrying modified target genes and detection or selection marker genes. The various elements useful for such strategies, and how to employ them, are described below.

Target gene modification can be accomplished by disruption of a coding sequence by insertion or deletion of nucleotides or nucleotide stretches. Such insertions or deletions may be of any suitable size. Preferably, they are of a size of at least 2 nucleotides, for example up to 5, up to 10, up to 25 or up to 50 nucleotides in length, excepting deletions which are multiples of 3. Alternatively, the

coding region of the target gene may be replaced by that of a (marker) gene. This confers an easily detectable phenotype on cells transformed with such a construct. Suitable examples of replacement genes are lacZ, xylE, Green Fluorescent Protein, and genes for the biosynthesis of antibiotics, such as erythromycin, apramycin, hygromycin, and thiostrepton, and metabolite analogues, such as fluoroacetamide.

Transfer of a disrupted target gene to a Pimaricin production host, resulting in *in vivo* gene inactivation, can be accomplished by using e.g. suicide vector systems, a defective phage containing a fragment internal to the coding region of the target gene, or a variant of the gene inactivated through deletion or insertion of DNA stretches as described above, and optionally a detection or selection marker. Suicide vectors and defective phages are characterized by their inability to propagate autonomously in the strain to be transformed and thus cannot be stably maintained by themselves. For *Streptomycetes* in general several suicide systems are available and suicide vectors can be chosen from the group of extrachromosomal element based cloning vectors available for *E. coli*, which cannot replicate in *Streptomyces* species, including for example pBR322, pUC, CoID, RSF1010, RK2 and vectors derived from these plasmids. Similarly, *Streptomyces* plasmids characterized by a limited host range can be selected that are incapable of stable maintenance in the desired host strain. Examples of such narrow host range plasmids are SLP1.2 and SCP2, and cloning vectors derived from these plamids. Still another possibility is to use temperature sensitive variants of *Streptomyces* wide host range plasmids. These plamids are characterized by their inability to replicate above a certain (restrictive) temperature. Besides non-replicative plasmids, defective phage vectors have been developed based on the *Streptomyces* phage phiC31 and have proven extremely useful for genetic analysis. In this

regard, it is noted that an extensive overview of known *Streptomyces* genetic engineering techniques may be found in Hopwood et al. (D.A. Hopwood, M.J. Bibb, K.F. Chater, T. Kieser, C.J. Bruton, H.M. Kieser, D.J. Lydiate, C.P. Smith, J.M. Ward, H. Schrempf, Genetic Manipulation of *Streptomyces*: A Laboratory Manual, The John Innes Foundation, Norwich, England, 1985).

The above mentioned suicide constructs can be introduced in a desired host cell using transformation procedures with isolated DNA, by conjugation from a donor microorganism, e.g. an *E. coli* or *Streptomyces* strain harboring the construct, or via transfection by phage particles. All of these methods are well within the knowledge of the person skilled in the art.

Upon introduction of such a construct in the microorganism of interest, e.g. *Streptomyces natalensis*, stable maintenance of the introduced genetic information is only possible by integration of the construct in the host chromosome, preferably by homologous recombination with the chromosomal copy of the target gene. Strains having integrated the construct in the chromosome can be detected by the expression of a co-introduced marker. In case of a detection marker, transformed colonies can be screened for acquired properties such as conversion of a colorless substrate into a colored compound (applicable with e.g. the genes lacZ, or xylE) or fluorescence (by expression of e.g. Green Fluorescent Protein). Alternatively, a marker can be used which allows selection of transformed strains by acquired resistance to e.g. antibiotics or toxic metabolite analogues. The latter method usually is employed more frequently because only cells with the acquired resistance will be able to grow in media containing the antibiotic or toxic metabolite analogue. If an internal fragment of the target gene is used for the construction of the suicide vector or defective phage, integration of the construct into the chromosomal copy of the target gene will result in

inactivation immediately. If the suicide construct or defective phage contains the complete target gene or a fragment including the N-terminal or C-terminal coding region, though inactivated through smaller insertions or deletions, only integration of the construct will result in the presence of an active and inactive copy of the gene, separated by vector DNA. For obtaining a strain with only an inactive copy, a second homologous recombination has to take place removing the vector sequences and the active copy of the target gene. Strains having undergone this second homologous recombination can be detected by the loss of the acquired property encoded by the co-introduced marker gene.

Another application of the present genes from the Pimaricin gene cluster lies in overexpression of one or more of these genes in the natural host, *Streptomyces natalensis*. The expression of the individual genes within the cluster is tightly regulated by the cell physiology and/or cluster specific regulatory genes. This internal control may be appropriate for production of the antibiotic in the natural environment, but is undesirable for industrial production. Overexpression of all genes of the cluster by introduction of additional gene copies or altering the controlling elements (e.g. promoters or regulatory genes) can boost antibiotic production considerably. This has been shown for e.g. Actinorhodin production by *Streptomyces coelicolor*. A similar effect can be obtained by overexpression, specifically of those genes encoding enzymes representing rate limiting steps in antibiotic biosynthesis.

Additional copies of each of the present genes from the Pimaricin biosynthesis gene cluster or homologues or fragments thereof, either separately or in different combinations, can be introduced into *Streptomyces natalensis*. This increases the efficiency of the oxidative reactions leading to biosynthesis of the natural Pimaricin molecule, and results in strains displaying improved Pimaricin production. This increase may be in the form of

increased Pimaricin titre in the culture broth or a higher product yield on substrate consumed. Of course, enhanced expression of certain genes can also be combined with inactivation of other genes, thus effecting improved production of variants of Pimaricin as described above.

Strains containing additional copies of target genes can be obtained through introduction of complete genes including expression signals (promoters and optionally enhancers) into the host chromosome. Suitable techniques include suicide vectors and defective phage, as described above. Alternatively, autonomously replicating DNA molecules derived from phage genomes or extrachromosomal elements, for example plasmids, can be used to carry the additional genes. Suitable cloning vectors include those derived from plasmids pIJ101 and SCP2. Other vectors can be constructed based on the plasmid naturally occurring in *Streptomyces natalensis*, as disclosed in GB patent application nr 2210619, using selection and/or detection markers similar to those employed for the pIJ101 derived vectors, such as pIJ702, pIJ486, with or without added markers as described above.

For gene expression, a large variety of promoters efficiently directing transcription of genes in Streptomyces is available. An example of a constitutive promoter is the ermE promoter, which directs expression of the erythromycin resistance gene from *Saccharopolyspora erythraea*. By contrast the agarase gene promoter from *S.coelicolor*, the promoter of the glycerol utilization operon, or the tipA promoter are examples of promoters inducible by specific substrates. Using techniques known in the art additional promoters can be obtained, e.g. promoters endogenous to *S.natalensis* (see J.M.Ward, G.R.Janssen, T.Kieser, M.J.Bibb, M.J.Buttner, M.J.Bibb. 1986. Mol.Gen.Genet. 203: 468-478).

The degree of overexpression can be manipulated by the choice of the promoter, by the amount of inducing compound, or by the choice of the autonomously replicating vector systems. Depending on the vector derivative used,

predetermined plasmid copy numbers can range from 1 or 2 to about 500. It is well within the expertise of the normal person skilled in the art to adjust the vector system to the desired degree of overexpression.

Both of the above uses of polynucleotides of the invention, i.e. inactivation to obtain new variants of Pimaricin and overexpression to increase Pimaricin productivity, can also be applied to strains producing structurally similar bioactive compounds for instance polymer antibiotics such as Amphotericin B (*Streptomyces nodosus*), Nystatin (*Streptomyces noursei*)(see Figure 4) to obtain variants of these compounds and/or to improve productivity Using the present genes to inactivate the corresponding genes in *Streptomyces* species other than *Streptomyces natalensis* will result in new derivatives of, *inter alia*, nystatin and amphotericin B which are altered in their oxidative state.

A further application of the polynucleotides of the invention is the heterologous expression and exploitation of the enzymatic activity encoded by one or more of the said polynucleotides. Using similar vector systems as employed for overexpression of the oxidative genes in *S.natalensis*, other microorganisms, preferably *Streptomycetes* species for instance the strain *Streptomyces lividans* or *Streptomyces coelicolor*, can be genetically transformed and thus acquire new oxidative enzymatic activity. This route is particularly useful for application of the enzymatic activities of polypeptides of the invention to the oxidative modification of other, preferable bioactive, compounds. Examples include secondary metabolites, antibiotics and anticancer agents etc., which often are highly functionalized chemical entities. Thus, it is possible to introduce one or more of the polynucleotides of the invention into a host producing such bioactive compounds naturally, or one which has acquired the genetic information to produce compounds by recombinant DNA technology. A strain having acquired a gene

or genes encoding oxidative enzymatic activity from the Pimarcin biosynthetic gene cluster will then be able to introduce, for example, epoxide functions or alcohol, aldehyde, or carboxyl groups into metabolites previously not modified in such a way. In this way it is possible to oxidize a methyl group which is not part of an linear alkane. A methyl group forming part of an aliphatic ring of an organic compound or biocompound can be oxidized by one or more of the polypeptides of the invention. The polypeptides of the invention can be isolated or purified from rDNA transformed hosts in which one or more of the polynucleotides of the invention are introduced. Preferably the polynucleotide are heterologous to the host. But also the transformed host as such may be used for the oxidative conversion. Thus, an approach has been provided, which allows for the creation of new variants of bioactive compounds not obtainable by chemical means (exemplified in Example 6 below).

The invention will now be demonstrated by the following, non-restrictive examples.

Examples

Example 1. Isolation and identification of Pimaricin biosynthetic genes.

Streptomyces natalensis strain ATCC27448 was grown in YEME medium (D.A. Hopwood, M.J. Bibb, K.F. Chater, T. Kieser, C.J. Bruton, H.M. Kieser, D.J. Lydiate, C.P. Smith, J.M. Ward, H. Schrempf, Genetic Manipulation of *Streptomyces*: A Laboratory Manual, The John Innes Foundation, Norwich, England, 1985) at 30°C for 3 days. Mycelium was harvested and total DNA was extracted and purified essentially as described by Hopwood (ibid.).

Total *S.natalensis* DNA was subjected to partial digestion with the restriction enzyme *Sau*3AI and size

fractionated on 0.8% agarose gel. Fragments of 30-40 kbp were isolated, inserted into *Bam*HI digested cosmid Supercos1 and subsequently introduced in *E.coli* strain XL1-Blue MR according to protocols suggested by the supplier (Stratagene, La Jolla).

Thus, a cosmid library of *S.natalensis* DNA in *E.coli* was obtained. The cosmid library was screened for the presence of polyketide synthase (PKS) related sequences by hybridization with radioactively labeled fragments from known PKS genes from the Rapamycin biosynthesis cluster from *Streptomyces hygroscopicus* (T.Schwecke, J.F.Aparicio, Y.Molnár, A.König, L.E.Khaw, S.F.Haydock, M.Oliynyk, P.Caffrey, J.Cortés, J.B.Lester, G.A.Böhm, J.Staunton, P.F.Leadlay. 1995. Proc. Natl. Acad. Sci. USA 92: 7839-7843).

Several clones were isolated which contained sequences hybridizing to a fragment containing the KS module 5 of rapB.

Complete DNA sequence determination of a number of neighbouring *Not*I fragments from Cos9 was performed after cloning the fragments in pBluescript. Computer assisted analysis of the DNA sequences revealed the presence of genes clearly identifiable as PKS gene modules on the basis of nucleotide and derived amino acid sequence homology with established PKS genes and proteins involved in the biosynthesis of erythromycin and rapamycin, as well as with fatty acid synthase genes and proteins, which catalyze a similar set of reactions. The complete nucleotide sequences and derived amino acid sequences of two Pimaricin PKS genes are given as SEQ ID numbers 1-4.

Using a 0.7 kb *Not*I fragment from Cos9 as a probe, the extent of the PKS related genes on the cosmid map was established as indicated in Figure 1.

Example 2. PKS genes are essential for Pimaricin biosynthesis

A completely sequenced 3.3 kb *Not*I DNA fragment (see Figure 1)(in pBluescript), encoding part of a *S.natalensis* PKS as deduced form the organizational and structural sequence similarities with known PKS, was excised by *Sac*I from the sequencing vector. The fragment was subcloned into the phage vector KC515 (M.R.Rodicio, C.J.Bruton, K.F.Chater. 1985. Gene 34: 283-292) and introduced in *S.lividans* to obtain infectious particles (recombinant phage) containing the *S.natalensis* PKS fragment. Infection of *S.natalensis* using this recombinant phage population and selection for resistance to the antibiotic viomycin, allowed the isolation of lysogens, originated through integration of the recombinant phage DNA into the *S.natalensis* chromosomal DNA by homologous recombination of the PKS regions.

None of 20 lysogens tested displayed antifungal activity as analyzed by an agar plate bioassay using *Candida utilis* as the indicator organism. Detailed analysis of one of the lysogens by Southern hybridization studies confirmed that integration of the recombinant phage DNA into the *S.natalensis* chromosomal PKS locus had indeed occurred.

Culturing the lysogen with the disrupted PKS gene in standard production medium (25 g/l soya peptone, 0.5 mM $ZnSO_4$, 20 g/l glucose, pH 7.5) followed by extraction of the culture broth with butanol, and UV spectrophotometric analysis indicated that no traces of Pimaricin were produced by this lysogen (J.F.Martín, A.L.Demain. 1975. Biochem. Biophys. Res. Commun. 71: 1103-1109).

Example 3. Detailed sequence analysis of non-PKS genes; preliminary identification.

Full sequence analysis of the regions flanking the PKS genes of Example 1 revealed the presence of additional

open reading frames (ORF) potentially encoding proteins functional in Pimaricin biosynthesis.

Homology comparison of the deduced amino acids sequences of the ORFs indicated the involvement of several in oxidation/reduction reactions. ORF1 showed a clear homology with previously identified cholesterol oxidases and ORF2 and ORF3 were similar to cytochrome P-450 monooxygenase proteins. Also, genes encoding accessory proteins for the P-450 enzymes seem to be present i.e. ferredoxin type. Complete nucleotide sequences of the respective genes and derived amino acid sequences are added as SEQ ID numbers 5-10. Detailed information on the chromosomal regions enompassing the three open reading frames (ORF's) is presented in Figure 2.

Example 4. Functional characterization of non-PKS genes involved in Pimaricin biosynthesis.

To define the involvement of the accessory genes/proteins in Pimaricin biosynthesis, both ORF1 and ORF3 were disrupted and the effect on Pimaricin production established. Similar strategies as described in Example 2 for the PKS disruption were employed for the non-PKS genes. ORF1 : a 7kb *Sph*I fragment containing the complete ORF1 was cloned into pUC19, the resulting plasmid was digested with *Bgl*II, the cohesive ends were filled in by treatment with Klenow polymerase and religated. This new plasmid was used as a source for DNA for the gene replacement. The 2.9 kb *Bam*HI-*Pst*I fragment from the plasmid was cloned into the *Bam*HI-*Pst*I sites of KC515. The recombinant phage was propagated in *S.lividans*, and used to infect the wildtype *S.natalensis* strain. Lysogens were obtained by selection for thiostrepton. The second recombination event was searched for by the loss of thiostrepton resistance. The insertion and subsequent loss of the phage as well as the final structure of the disruptred gene was confirmed by Southern

hybridization.

ORF3 : disruption was accomplished by insertion of a 667 bp *Pvu*II-*Sma*I fragment internal to ORF3 in *Hin*CII cut pUC19; The fragment was excised using *Bam*HI and *Pst*I and ligated into similarly digested phage vector KC515. Transformation of the ligation mixture to *S.lividans* yielded recombinant phage Ø6D4-1particles. After transfection of *S.natalensis*, lysogens were isolated as described above. Disruption of ORF3 in *S.natalensis* mutant D4 was confirmed by Southern hybridization

Example 5. Analysis of ORF1 and ORF3 gene disruptants of *S.natalensis*.

Strains with disrupted ORF1 and ORF3 were analyzed for pimaricin production using the bioassay with *C.utilis*. For both disruptants the production of an antifungal activity was strongly reduced as compared with the wild-type strain *S.natalensis* ATCC27448.
Both strains were cultured in pimaricin production medium (see Example 2) and the culture filtrate was analyzed by combined liquid chromotography/mass spectroscopy (LC-MS) analysis.

Disruptants in ORF1 did not contain any pimaricin-like molecule in the culture filtrate.
In the case of the ORF3 disruptant a single Pimaricin-like molecule was detected in the culture filtrate having molecular mass of 649.75 indicating the loss of exactly 1 oxygen atom. The exact structure was determined by NMR spectroscopy to be identical to Pimaricin except that the epoxide function at was replaced by a double bond; the structure with a double bond between C_4-C_5 (displayed in Figure 3b (top)) is the expected biosynthetic precursor for the epoxidation.

Example 6. Overexpression of ORF1, ORF2, and ORF3 in *S.natalensis*.

Separate overexpression of ORF1, ORF2 and ORF3 was achieved by placing each gene under the direction of the ermE promoter from *Saccharopolyspora erythraea* (M.J. Bibb, G.R. Janssen, J.M. Ward. 1985. Gene 38: 215-226). A useful derivative of this promoter, having a number of cloning sites attached was obtained by PCR using the following oligonucleotides: SEQ ID 11:
AAACTGCAGCTCTAGAGGCGGCTTGCGCCCGATGCTAGTC

SEQ ID 12:
AAACTGCAGCTCTAGATGCCCGGGTATCGATCGTCGACGGCATGCGGATCCTACCAACCG GCACGATTG

The 225 bp PCR fragment obtained was digested with *Pst*I, purified by agarose gel electrophoresis and inserted into PstI digested pUC19, yielding pUCermE

ORF1 was inserted in pUCermE as a 2.2 kb *Sph*I-*Cla*I fragment encompassing the complete coding sequence. For ORF2 a 3.5 kb *Cla*I-*Nru*I fragment was used, and for ORF3 a 2.8 kb *Sal*I-*Kpn*I fragment was used. Each ermE promoter-ORF combination was subsequently excised as a *Pst*I fragment, inserted in *Pst*I digested phage vector KC515 and introduced in S.natalensis essentially as described in Example 4.

Recombinant *S.natalensis* strains were thus obtained which overexpressed one of the three genes. Each strain showed improved levels of Pimaricin production of 10 -15 % after growth under standard production conditions (see Example 2).

Example 7. Expression of *S.natalensis* ORF1, ORF2, and ORF3 in *S.coelicolor* and *S.lividans*

ORF1 and ORF2: A 223 bp *Nde*I-*Eco*RI fragment,

corresponding to the 5´end of ORF1 from the ATG to the first *Eco*RI site was obtained using the Polymerase Chain Reaction such that an *Nde*I site was created coinciding with the ATG initiation codon of ORF1. The oligonuleotides used for this PCR were 5'-AGGATTACCCATATGTTCGAGAACCAGCAT-3' (forward; SEQ ID NO 13) and 5'-GCATGAGCGTGGGAATTCCG -3" (reverse; SEQ ID NO 14). The PCR product was digested with *Nde*1 and *Eco*RI cloned into similarly digested vector pT7-7 (S. Tabor, C.C. Richardson. 1985. PNAS 82, 1074) to yield plasmid pJA56.

pJA56 was digested with *Eco*RI and *Sma*I, and ligated to an *Eco*RI-*Nru*I fragment encompassing ORF1 and ORF2, yielding plasmid pJA57.

pJA57 was digested with *Nde*I and ligated to *Nde*I-digested pIJ6021 (E .Takano et al. 1995. Gene 166,133).The resulting plasmid was named pJA58. Both ORF1 and ORF2 are now under the direction of the thiostrepton inducible *tip*A promoter. Plasmid pJA58 was transformed into strain *S.coelicolor* A(3)2 and *S.lividans* 1326.

ORF3 : The ORF3 expression vector has been constructed by cloning a 3.7 kb *Kpn*I fragment containing the complete ORF3 into the unique *Kpn*I site of pHZ1351 (Bao et al.. 1997. ISBA Meeting abstract 4P15). The resulting plasmid (pJA50) was transformed to strain *S.coelicolor* A(3)2 and *S.lividans* 1326. Expression of ORF3 is directed by its own promoter.

Example 8. Activity of cell-free extracts of *S.coelicolor* expressing ORF1, ORF2, and ORF3.

S.coelicolor strains expressing the genes ORF1 and 2, and ORF3, respectively, thus producing the active proteins pORF1, pORF2, and pORF3 were grown in YEME medium (Hopwood et al., ibid). For induction thiostrepton was added to 0.005mg/l. Incubation was for 48 hrs. at 30°C.

Cell-free extracts were prepared as follows: Mycelium was harvested by centrifugation at 5000xg/4°C for

10 minutes and washed with 1 volume of 50mM Tris-HCl pH 7.5, 1mM DTT, 10% glycerol. The mycelium was resuspended in 0.2 volume of 50mM Tris-HCl pH 7.5, 1mM DTT, 10% glycerol; 1 tablet of protease inhibitor cocktail (Boehringer Mannheim) was added per 25 ml of extract. Cell extracts were prepared by sonication. After sonication cell debris were removed by centrifugation at 10000xg / 4°C for 10 minutes.

Activity assays for the cell-free extracts were performed using *S.coelicolor* cell-free extract (100-1000μg total protein); 0.5 μmol NADPH; 5 μmol glucose-6-phosphate; 0.5 U glucose-6-phosphate dehydrogenase; 22μg spinach ferredoxin; 0.05 U spinach ferredoxin NADP+ reductase. As substrate for the oxidation activities triketide lactone (TKL, see Figure 5; M.J.B. Brown et al. 1995. J.Chem.Soc. Chem.Comm. 1517; C.M. Kao et al.. 1995. J.Am.Chem.Soc. 117, 9105) was added. After allowing to react for 60-90 minutes, the products were extracted twice with an equal volume of ethylacetate, and analysed by thin layer chromatography, LC-MS, and NMR spectroscopy.

It appeared that pORF3 was inactive on this specific substrate, but that the combined action of pORF1 and pORF2 resulted in a TKL derivative having the methyl group completely oxidized to the carboxylic acid function (see Figure 5).